DOCUMENT RESUME

ED 474 272 TM 034 807

AUTHOR	Schroeder, Carolyn M.
TITLE	There Are Infinitely Many Normal Distributions: Not All Normal Distributions Are Standard Normal.
PUB DATE	2003-02-00
NOTE	15p.; Paper presented at the Annual Meeting of the Southwest Educational Research Association (San Antonio, TX, February 13-15, 2003).
PUB TYPE	Reports - Descriptive (141) -- Speeches/Meeting Papers (150)
EDRS PRICE	EDRS Price MF01/PC01 Plus Postage.
DESCRIPTORS	*Statistical Distributions
IDENTIFIERS	Curves; *Normal Distribution; *Univariate Analysis

ABSTRACT

Many researchers and students assume that the univariate distribution has a single, ubiquitous "bell" shape, perhaps because most books only portray the "standard normal" or normal z-score distribution. This paper shows that the normal curve can have infinitely many appearances. It illustrates that the width to height ratio can change an infinite number of times as the mean and standard deviation change, making the curve appear slightly different each time, while remaining a normal curve. Trying to judge normality by simply looking at the curve may be deceptive. The final judgment must be made by checking the criteria for normality: symmetry, bell shape, unimodality, and fit to the equation for normality. (Contains 6 figures and 11 references.) (Author/SLD)

Reproductions supplied by EDRS are the best that can be made
from the original document.

ED 474 272

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY

C. M. Schroeder

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

1

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

There Are Infinitely Many Normal Distributions:

Not All Normal Distributions Are Standard Normal

Carolyn M. Schroeder

Texas A&M University

TM034807

Paper presented at the annual meeting of the Southwest Educational Research Association

San Antonio, TX Feb.13-15, 2003

Thanks to Dr. Bruce Thompson for the use of his unpublished data set in this paper.

BEST COPY AVAILABLE

Abstract

Surprisingly, many researchers and students assume that the univariate distribution has a single ubiquitous "bell" shape, perhaps because most books only portray the "standard normal" or normal z-score distribution. This paper will show that the normal curve can have infinitely many appearances.

There Are Infinitely Many Normal Distributions:

Not All Normal Distributions Are Standard Normal

Many univariate statistical methods assume that dependent variable data have a univariate normal distribution, and some statistical methods assume that the error scores are normally distributed (Thompson, 1992). Almost always books portray only a single case of the infinitely many univariate normal distributions: the "standard normal" or "normal z-score" distribution, thus implicitly teaching the misconception that the normal distribution "bell" takes a single shape. In reality, the normal curve can take infinitely many forms, each having skewness and kurtosis coefficients of zero (Bump, 1991; Burdenski, 2000; Henson, 1999), but differing in appearance (e.g., apparent width and apparent height).

A thorough understanding of normal distributions is necessary because of their frequent occurrence in nature as well as in testing, particularly when large samples are considered. Examples of data which yield normal distributions include the height and weight of adult human beings, IQ scores, chances of throwing a certain total score when throwing 2 or more dice, leaf length of mesquite trees, and standardized test scores. Combining the concepts of standard scores and normal distributions gives us a valuable tool for analyzing distributions of many kinds of variables (Hinkle, Wiersma, and Jurs, 1998).

A normal distribution is theoretical, symmetric, bell-shaped, unimodal, and fits the equation

$$Y = \frac{1}{\sigma\sqrt{2\pi}} e^{-(X-\mu)^2/2\sigma^2}$$

where Y is the height of the ordinate for any given value of X in the distribution, π is equal

to 3.1416, the ratio of the circumference of a circle to its diameter, and *e* is 2.7183, the base of the natural logarithm system. The mean of the distribution scores is represented by μ and the standard deviation of the score distribution is σ. This equation was developed by the French mathematician Abraham Demoivre (1667-1754) who based it on his observations of games of chance, using it to predict the probability of different outcomes (Hinkle, Wiersma, & Jurs, 1998).

Spatz (2001) reports that about the beginning of the 19th century, Carl Friedrich Gauss determined the mathematics of the curve, using it as a way to describe random error in astronomical observations. This "Gaussian" curve was recognized as an accurate portrayal of the results of random variation, and it began to be called the law of error. An early and influential advocate of the normal curve, Belgian Adolphe Quetelet demonstrated that many empirical biological and social measurements correspond with this theoretical curve. Florence Nightingale, who pioneered the use of statistics to improve health care, called Quetelet "the founder of 'the most important science in the whole earth'" (Spatz, 2001, p. 119). In his attempt to standardize statistical terminology at the end of the 19th century, Karl Pearson began to call it the normal distribution, probably because of the large number of types of measurements which fall into this pattern.

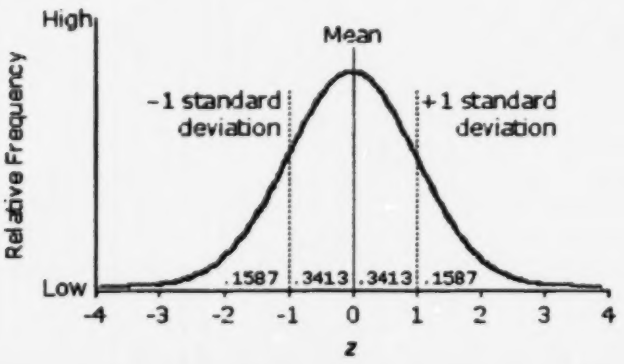

Figure 1 The Theoretical Normal Distribution (Lowery, 1999)

BEST COPY AVAILABLE

Normal distributions (See Fig. 1) are continuous and asymptotic to (never touch) the X axis. The area under the curve is equal to one, and the mean, median, and mode are equal to each other and correspond to the peak of the curve. Convention and convenience dictate that the curve is drawn from -3σ to +3σ from the mean. The steepest points of the curve, called inflection points, are located at exactly -1σ and +1σ, with approximately 68% of the scores lying between these points. Approximately 95% of the scores lie between -2σ and +2σ, and over 99% lie between -3σ and +3σ. The normal bell-shaped curve that is most familiar is in reality the "*standard normal curve* representing a standardized distribution of z-scores and is *not* the only normal curve" (Henson, 1999, p. 196). Hinkle et al. (1998), point out that because no specific mean or standard deviation is specified in the formula, every possible mean and standard deviation combination has its own individual normal distribution, resulting in what they call a "family of distributions" (p. 91). Because there is an infinite number of possible combinations of means and standard deviations, the number of possible normal distributions is also infinite. This symmetrical curve represents a situation where there are many scores near the center of the distribution and fewer and fewer scores as you move away from the center.

According to Bump (1991), normality of a curve is evaluated by using skewness and kurtosis in addition to mean and standard deviation. Mean, standard deviation, skewness, and kurtosis taken together are called the first four moments of a normal distribution. In the physical sciences, moment is a term describing the measurement of the tendency of a force to cause rotation around a point or axis of an object. When the forces pressing down on either side of the axis of the object, calculated for each side by the mass times the distance

from the axis, are equal, the object is in a balanced state. "In statistics, the term 'moment' denotes class frequencies that are analogous to the forces exerted in the previous example" (Bump, 1991, p. 3). In a histogram of a normally distributed data set, "the 'moment' contribution of each column is measured by the product of the class frequency (f) and the corresponding deviation (x) from the origin. The sum of the fx products, divided by the total frequencies, gives a net measure called the first moment or the mean" (p. 3).

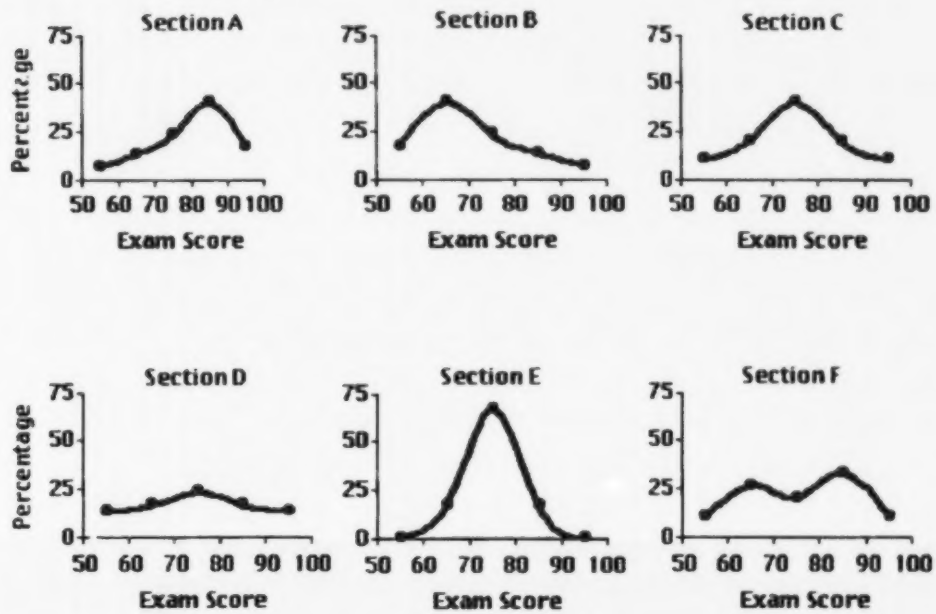

Figure 2 Distributions of Exam Scores in Six Sections of a Statistics Course (Lowery, 1999)

All normal distributions are symmetrical; however, not all symmetrical distributions are normal. Examples of non-normal symmetrical distributions may include bimodal distributions as in Section F in Fig. 2, rectangular uniform distributions, kurtic distributions as in Sections D and E and skewed distributions as in Sections A and B. Kurtosis refers to the degree of peakedness of the distribution, and distributions may be either leptokurtic

(extremely peaked, with most scores near the center), as in Fig. 2, Sec. E or platykurtic (flattened, with a more uniform distribution of scores) as in Fig. 2, Sec. D. A mesokurtic distribution, represented by section C, is more moderate than either the platykurtic or leptokurtic distribution. Kurtosis can be calculated using the formula

$$Coefficient\ of\ Kurtosis = \left[\frac{n(n+1)}{(n-1)(n-2)(n-3)}\right]\left[\sum\left(\frac{X_i - \overline{X}}{SD_X}\right)^4\right] - \left[\frac{3(n-1)^2}{(n-2)(n-3)}\right]$$

Distributions which have a large number of scores near one end or the other of the scale with fewer scores on either side are said to be skewed and are also not normal distributions (Hinkle et al., 1998). Section B above is said to be positively skewed while Section A is negatively skewed. Skewness can be calculated by the formula

$$Coefficient\ of\ Skewness = \left[\frac{n}{(n-1)(n-2)}\right]\left[\sum\left(\frac{X_i - \overline{X}}{SD_X}\right)^3\right]$$

Nonstandard normal distributions can be transformed into standard normal distributions (or normalized). This is particularly useful when interpreting standardized test results, such as those from achievement tests, IQ tests, and College Entrance Examinations. Transformed scores help to overcome some of the difficulties and disadvantages involved in reporting test results as *z* scores, including the negative connotations of receiving a negative score as perceived by students and parents, the possibility of interpreting a *z* score of zero as having answered no questions correctly, and the implication by the usual 2 decimal places of

a z score of a degree of precision possibly not possessed by the raw score. In order to transform a distribution, the raw scores are first converted to z scores (standard scores) and then changed into a distribution with a predetermined mean and standard deviation by the formula

$$X' = (s')(z) + \bar{X}'$$

where X' is equal to the score (new or transformed) of an individual, s' is equal to the predetermined standard deviation of the distribution, z is the individual's standard score, and $\bar{X}'$ is the predetermined mean of the distribution (Hinkle et al., 1998)

The heuristic illustration which follows presents a univariate data set (Thompson, 2003) with 100 cases that have a univariate normal distribution. Then various monotonic transformations are invoked to show that the distribution can take many forms. For each transformed distribution, summary statistics (i.e., the mean and moments about the mean, including standard deviation, skewness, and kurtosis) are computed, along with graphical overlays of the normal curve on histograms for each data set.

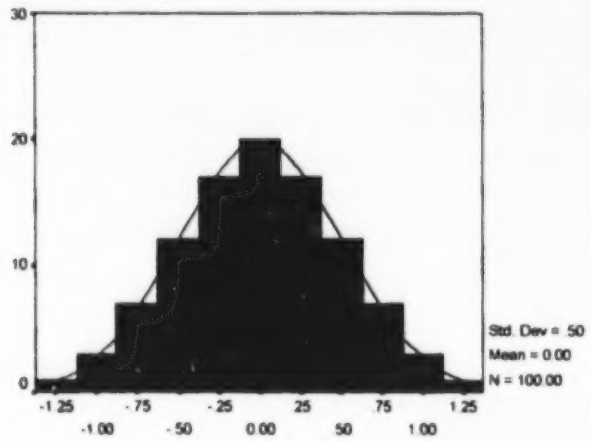

Figure 3 Histogram of the Original Data Set with Overlay of Normal Curve

The parameters of the sample data set included a mean of 0.0000 and a median of 0.000. The variance was 0.252, the standard deviation was 0.50247 and the scores had a range of 2.60, with a minimum of -1.30 and a maximum of 1.30. Skewness was 0.000, and kurtosis was -0.090. Figure 3 shows a histogram of the original data set overlaid by the normal curve, and reveals that the data fits the normal curve very well.

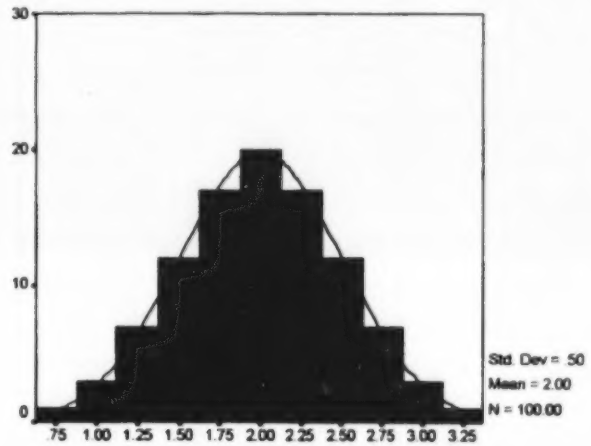

Figure 4 Histogram of the Additive Data Set with Overlay of the Normal Curve

BEST COPY AVAILABLE

The next step was to apply an additive constant of 2 to each score in the original data set and to determine its parameters. The new mean was 2.0000 and the new median was also 2.0000. The variance was 0.252, the standard deviation was 0.50247, and the scores had a range of 2.60, with a minimum of 0.70 and a maximum of 3.30. Skewness was 0.000 and kurtosis was -0.090. It can be seen, therefore, that the use of an additive constant changes the mean, median, and minimum and maximum scores. However, it does not affect the variance, standard deviation, range, skewness, or kurtosis. From the histogram in Figure 4 it can be seen that the only difference from the graph in Figure 3 is in the new numbers along the abscissa, which effectively move the new histogram to the right along the abscissa, but do not affect its size or shape.

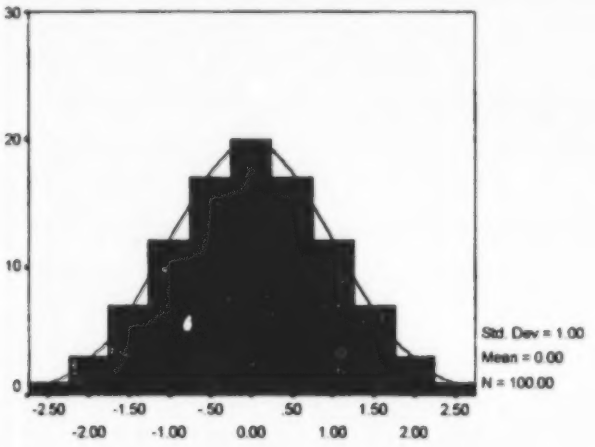

Figure 5 Histogram of the Multiplicative Data Set with Overlay of the Normal Curve

A multiplicative constant of 2 was then applied to each score in the original data set, resulting in a new mean and new median of 0.0000. The variance was 1.010, the standard deviation was 1.00494, and the range was 5.20 with a minimum of -2.6 and a maximum of 2.60. Again, skewness was 0.000 and kurtosis was -0.090. A multiplicative constant does not affect the skewness or kurtosis, but does affect the variance, standard

BEST COPY AVAILABLE

deviation, and range. The mean is affected by multiplicative constants if (and only if) the original mean is not zero. Figure 5 shows a histogram of the results of application of the multiplicative constant to the original data set. However, one should notice that the scale of the abscissa has been changed by SPSS, making the appearance of the histogram exactly like those in Figures 3 and 4. If the original scale (points increasing by 0.25 instead of 0.5) had been used, the curve would appear to be much wider in proportion to its height. However, it would remain a normal curve.

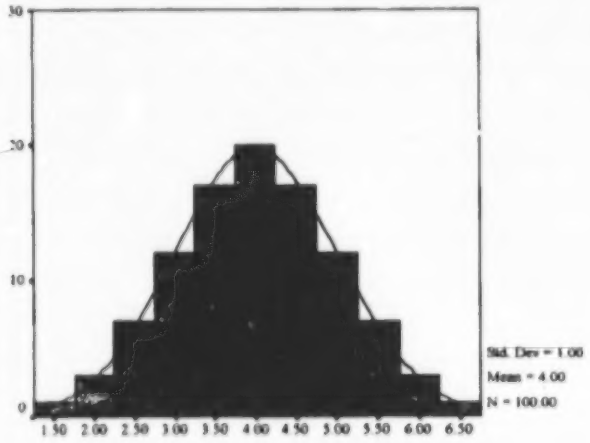

Figure 6 Histogram of Additive and Multiplicative Data Set with Overlay of the Normal Curve

Finally, an additive constant of 2 and then a multiplicative constant of 2 were applied to the original data set, resulting in a mean of 4.0000 and a median of 4.0000. The variance was 1.010, the standard deviation was 1.00494, and the range was 5.20 with a minimum of 1.40 and a maximum of 6.60. Skewness was still 0.000 and kurtosis was -0.090. Applying an additive constant and then a multiplicative constant resulted in changes from the original data set in mean, median, variance, standard deviation, and ranges. However, it did not affect skewness or kurtosis. As can be seen in Figure 6, the

BEST COPY AVAILABLE

appearance of the histogram did not change from Figure 5, although it is moved along the abscissa in a positive direction.

Summary

The concept of the normal curve is extremely important in statistics. Its mathematical properties are useful in understanding many distributions in nature, because often they are normally or near-normally distributed. It can be seen that the width to height ratio can change an infinite number of times as the mean and standard deviation change, making the curve appear slightly different each time, although it is still a normal curve. Care must be exercised in determining whether or not a distribution exhibits normality. Trying to judge normality by simply looking at the curve may be deceptive. The final judgment must be made by checking the criteria for normality: symmetry, bell shape, unimodality, and fit to the equation for normality.

References

Bump, W. (1991, January). *The normal curve takes many forms: A review of skewness and kurtosis.* Paper presented at the annual meeting of the Southwest Educational Research Association, San Antonio. (ERIC Document Reproduction Service No. ED 342790)

Burdenski, Jr., T.K. (2000, April). *Evaluating univariate, bivariate and multivariate normality using graphical procedures.* Paper presented at the annual meeting of the American Educational Research Association, New Orleans. (ERIC Document Reproduction Service No. ED 440989)

Henson, R.K. (1999). Multivariate normality: What is it and how is it assessed? In B. Thompson (Ed.), *Advances in social science methodology* (Vol. 5, pp. 193-212). Stamford, CT: JAI Press

Hinkle, D.E., Wiersma, W., & Jurs, S.G. (1998). *Applied statistics for the behavioral sciences* (4th ed.). Boston: Houghton Mifflin.

Lowery, R. (1999). *Concepts and Applications of Inferential Statistics.* Retrieved January 3, 2003, from http://vassun.vassar.edu/%7Elowry/webtext.html

Mendenhall, W. (1983). *Introduction to probability and statistics* (6th ed.). Boston: Duxbury Press.

NIST/SEMATECH e-Handbook of Statistical Methods. (n.d.). Retrieved February 5, 2003 from http://www.itl.nist.gov/div898/handbook/eda/section3/eda35b.htm

Spatz, C. (2001). *Basic statistics* (7th ed.). Belmont, CA: Wadsworth/Thomson Learning.

Thompson, B. (1992, April). *Interpreting regression results: Beta weights and structure coefficients are both important.* Paper presented at the annual meeting of the American

Educational Research Association, San Francisco. (ERIC Document Reproduction Service No. ED 344897)

Thompson, B. (2003). Unpublished data set.

Wright, J. (2001). *Standard normal distribution.* Retrieved December 19, 2002, from http://www2.sunysuffolk.edu/wrightj/MA23/Chapter7

U.S. Department of Education
Office of Educational Research and Improvement (OERI)
National Library of Education (NLE)
Educational Resources Information Center (ERIC)

TM034807

REPRODUCTION RELEASE

(Specific Document)

I. DOCUMENT IDENTIFICATION:

Title: THERE ARE INFINITELY MANY NORMAL DISTRIBUTIONS: NOT ALL NORMAL DISTRIBUTIONS ARE STANDARD NORMAL	
Author(s): CAROLYN M. SCHROEDER	
Corporate Source:	Publication Date: 2/14/03

II. REPRODUCTION RELEASE:

In order to disseminate as widely as possible timely and significant materials of interest to the educational community, documents announced in the monthly abstract journal of the ERIC system, *Resources in Education* (RIE), are usually made available to users in microfiche, reproduced paper copy, and electronic media, and sold through the ERIC Document Reproduction Service (EDRS). Credit is given to the source of each document, and, if reproduction release is granted, one of the following notices is affixed to the document.

If permission is granted to reproduce and disseminate the identified document, please CHECK ONE of the following three options and sign at the bottom of the page.

The sample sticker shown below will be affixed to all Level 1 documents	The sample sticker shown below will be affixed to all Level 2A documents	The sample sticker shown below will be affixed to all Level 2B documents
PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY Sample TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) 1	PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE, AND IN ELECTRONIC MEDIA FOR ERIC COLLECTION SUBSCRIBERS ONLY, HAS BEEN GRANTED BY Sample TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) 2A	PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE ONLY HAS BEEN GRANTED BY Sample TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) 2B
Level 1 ↑ XX	Level 2A ↑ ☐	Level 2B ↑ ☐
Check here for Level 1 release, permitting reproduction and dissemination in microfiche or other ERIC archival media (e.g., electronic) *and* paper copy.	Check here for Level 2A release, permitting reproduction and dissemination in microfiche and in electronic media for ERIC archival collection subscribers only	Check here for Level 2B release, permitting reproduction and dissemination in microfiche only

Documents will be processed as indicated provided reproduction quality permits.
If permission to reproduce is granted, but no box is checked, documents will be processed at Level 1.

I hereby grant to the Educational Resources Information Center (ERIC) nonexclusive permission to reproduce and disseminate this document as indicated above. Reproduction from the ERIC microfiche or electronic media by persons other than ERIC employees and its system contractors requires permission from the copyright holder. Exception is made for non-profit reproduction by libraries and other service agencies to satisfy information needs of educators in response to discrete inquiries.

Sign here, → please

Signature: Carolyn Schroeder	Printed Name/Position/Title: CAROLYN M. SCHROEDER
Organization/Address: COLLEGE OF ED -- TLAC COLLEGE STATION, TX 77843-4232	Telephone: 979/845-1335 FAX: E-Mail Address: Date: 3/4/03

(over)

4232

III. DOCUMENT AVAILABILITY INFORMATION (FROM NON-ERIC SOURCE):

If permission to reproduce is not granted to ERIC, *or*, if you wish ERIC to cite the availability of the document from another source, please provide the following information regarding the availability of the document. (ERIC will not announce a document unless it is publicly available, and a dependable source can be specified. Contributors should also be aware that ERIC selection criteria are significantly more stringent for documents that cannot be made available through EDRS.)

Publisher/Distributor:
Address:
Price:

IV. REFERRAL OF ERIC TO COPYRIGHT/REPRODUCTION RIGHTS HOLDER:

If the right to grant this reproduction release is held by someone other than the addressee, please provide the appropriate name and address:

Name:
Address:

V. WHERE TO SEND THIS FORM:

Send this form to the following ERIC Clearinghouse:

University of Maryland
ERIC Clearinghouse on Assessment and Evaluation
1129 Shriver Laboratory
College Park, MD 20742
Attn: Acquisitions

However, if solicited by the ERIC Facility, or if making an unsolicited contribution to ERIC, return this form (and the document being contributed) to:

ERIC Processing and Reference Facility
1100 West Street, 2nd Floor
Laurel, Maryland 20707-3598

Telephone: 301-497-4080
Toll Free: 800-799-3742
FAX: 301-953-0263
e-mail: ericfac@inet.ed.gov
WWW: http://ericfac.piccard.csc.com

EFF-088 (Rev. 9/97)
PREVIOUS VERSIONS OF THIS FORM ARE OBSOLETE.